Introduction:

In this analysis, I compared proportions of adverse event reports of the Recovery, G2, G2x, G2 and G2x combined, and Eclipse vena cava filters relative to their sales, to the proportions of adverse event reports for the SNF VCF. I considered six time periods, which correspond to existing BARD AE reports: 2000- Q2'03, 2000-Q3'04, 2000-Q3'05, 2000-11/2007, 2000-11/2009, 2000-7/2010. In separate analyses, I considered data from 2000-April 23, 2004, which was of special interest due to hold on the Recovery filter that was imposed in the first quarter of 2004, from 2000-February 9, 2006 and from 2000-June 30, 2006.

The above analysis (and discussion below) is identical to my August, 28, 2016 report. Since that time I have conducted an additional analysis based on additional sales and adverse event data, and including the Denali and Meridian devices. I considered fracture reports for an additional six time periods: 2000-2009, 2000-2010, 2000-2011, 2000-2012, 2000-2013, and 2000-2014. In a separate analysis, I considered data from 2000-May, 2011, because Bard did an internal comparison using this time-period.

Data sources:

Data for this analysis are comprised of adverse event reports and monthly sales totals. The adverse event reports were extracted from the MAUDE database that is maintained by the FDA for the purpose of reporting for medical devices, as well as Trackwise (Q2'03: BPVE-01-00196343-maude2000.xls, April 23, 2004: BPVE-01-00268632, Q3'04: BPVE-01-00052935-maude.xls, Q3'05: BPVE-01-01054793.xlsx and BPV-17-01-0193291.xls, 11/2007: BPV-17-01-00188520.xls, 11/2009: BPVE-01-01501003.xls, 7/2010: BPVEFILTER-01-00050487.XLSX for BARD events and BPVEFILTER-01-00174270_2009 AERs.xls, BPVE-01-01706342_January 10 AEs.XLS, BPVEFILTER-01-00043057_July 10 2010.XLS, BPVE-01-00749928_February 10 AEs.XLS, BPVEFILTER-01-00043446_May 10 AEs.XLS, BPVEFILTER-01-00043059_March 10 AEs.XLS, BPVEFILTER-01-00043053_April 10 AEs.XLS, BPVEFILTER-01-00043058_June 10 AEs.XLS for SNF events). The sales data were provided by BARD (BPV-17-01-00193291.xlsx for sales from 2003 and BPV-17-01-00188520.XLS for sales prior to 2003).

For the updated analysis, I used the same sources of sales data, with the exception of the analysis through May, 2011, which contained both sales and adverse event data in BPVEFILTER-01-00037664. The adverse event reports were found in the following documents: 2008: BPV-17-01-00170706, 2009: BPV-17-01-00170503, 2010, 2011 and 2011: BPV-17-01-00170625, 2013: BPV-17-01-00226338, 2014: BPVEFILTER-01-00303182.

Data inconsistencies and errors:

I note that there was an error in the calculation of migration AE's on the Recovery tab of BPV-17-01-00188520.xls: [redacted] I also note that on the SNF tab of that same sheet, [redacted] In another sheet (BPVE-01-01501003), the [redacted]

[redacted] In the sheet that reports adverse events through July 2010 (BPVEFILTER-01-00050487 - Marauder Report by Manufacturer Summary IVC Filters 10-14-10 v5), [redacted] In contrast, in the sheet that reports adverse events through November, 2009 (BPVE-01-01501003), [redacted]

I have found substantial discrepancies between the company reports that I used for my analysis and internal company reports on Recovery adverse events. For example, in the data report that I used for my report through April 23, 2004 (BPVE-01-00268632), [redacted] In the data reports that I used through September, 2004 (BPVE-01-00052935-maude), [redacted] In a Bard memo dated September 7, 2004 (BPVE-01-01059656), [redacted] In the data reports that I used for my report through September, 2005 (BPVE-01-01054793), [redacted] In a Bard report titled [redacted] BPVE-01-00436350)," [redacted] These examples indicate that the data reports that I used provided considerably lower numbers of events for Recovery than originally detected by Bard.

The data used for the updated analysis also contain inconsistencies that are relatively minor and unlikely to affect the results. For example, while the data sheets refer to [redacted] (BPV-17-01-00170706, p. 5), while the [redacted] (BPV-17-01-00170503, p. 5). Where a discrepancy was noted, I used [redacted] with the later date, under the assumption that data errors may have been found and corrected over time.

The spreadsheet Bard produced with data cumulative through May, 2011 appears to have a different definition of [redacted] (BPVEFILTER-01-00037664), while data from earlier time-periods report [redacted] e.g. [redacted] (BPVE-01-01501003)). The numbers of [redacted]

Adverse events:

I considered five adverse events: deaths due to filter embolization, filter fracture, migration, perforation and tilt. In some instances of the extracted data in the existing spreadsheets the AE of [redacted]

[redacted] I considered both the original categories, as well as these enlarged categories. In later years, [redacted]

Statistical methods:

Reporting Risk Ratio (RRR) to compare AE's among products:

For a given time period, for each product and adverse event (AE), I calculated the reporting AE risk as the number of AE's divided by the total sales in that time period. I then calculated the reporting risk ratio (RRR) as the ratio of the reporting risk of each product to that of SNF. The reporting risk ratio is an estimate of the measure of interest, which is the risk ratio (RR). I discuss below (under Potential Limitations) conditions under which the RRR provides an unbiased or conservative estimate of the RR. Letting x_1 denote the number of AE's for the product of interest, x_2 denote the number of AE's for SNF, n_1 denote the total units sold for the product of interest and n_2 denote the total units sold for SNF, the RRR is defined as

$$RRR=(x_1/n_1)/(x_2/n_2).$$

A value of the RRR that is larger than one reflects a higher risk for the AE for the product of interest than for SNF. An RRR that is less than one reflects a lower risk of the AE for the product of interest than for SNF. Note that if there are 0 events for SNF (i.e., $x_2=0$), the RRR will involve division by 0. I have listed these instances of the RRR as "∞" in my tables. If there are 0 events for Recovery and 0 events for SNF, the RRR involves 0's in both the numerator and the denominator. I have listed these instances of the RRR as "0/0" in my tables. For example, considering data through Q3'05, there were 79,349 SNF units sold and 33,592 Recovery units sold. There were [redacted] higher risk for perforation associated with Recovery than with SNF.

I note that the reporting risk ratio is different from the proportional reporting ratio (PRR) and the reporting odds ratio (ROR), both of which use total numbers who were implanted with the devices who report AE's as denominators and not total numbers who were implanted with the devices.

p-values for inference about RRR

To test whether an observed RRR is statistically significantly different from one (i.e., the "null" value), I calculated the p-value. The p-value is the probability that the observed RRR [redacted] or an even more extreme RRR [redacted] could have arisen if the true RRR is actually one. If the p-value is very small (e.g., less than 0.05), we either have to believe that a highly unlikely event occurred, or that our presumption that the true RRR is equal to one is incorrect. Since the second possibility is more likely, we accept that explanation, and reject the null hypothesis that the true RRR is equal to one. That is, we conclude that the observed [redacted] is indeed significantly different from 1.

I calculated the p-values using two methods. For the first, I used approximate calculations that rely on large numbers. This approach approximates the natural log of the RRR divided by its approximate standard error as a normally distributed random variable and calculates the p-value on the basis of this approximation. The formula for the approximate standard error of the log(RRR) is given by:

$$\sqrt{\frac{1}{x1}+\frac{1}{x2}-\frac{1}{n1}-\frac{1}{n2}}$$

Where x_1 is the number of AE's for the product of interest (e.g., Recovery), x_2 is the number of AE's for SNF, n_1 is the total units sold for the product of interest and n_2 is the total units sold for SNF. Note that if there are 0 events for Recovery (i.e., $x_1=0$) or for SNF (i.e., $x_2=0$), this standard error cannot be calculated. In these instances I listed the p-value as "NA."

The second method calculates the p-value for the test of association between sales and AE's and is an "exact" test, meaning that it does not rely on any large sample approximations. I used Fisher's exact test to calculate these p-values. This approach is useful when there are zero AE's, as in many cases an exact p-value can be calculated, while it cannot for the approximate approach.

95% confidence intervals for the reporting risk ratio

I have calculated 95% confidence intervals for the reporting risk ratios for all of the analyses that include data through July 2010. The associated interpretation is that we can be 95% confident that the intervals contain the "true" reporting risk ratio. The lower bound of the interval is of greatest interest, because its distance from the value 1 is informative about the strength of the evidence in the data against a true reporting risk ratio of 1. The greater it is relative to 1, the smaller the associated p-value.

Adjustment for multiple testing

The chances of false positive findings increase with the number of statistical tests conducted. In the setting of analyses of efficacy, it is critical to account for this through some adjustment for the multiple testing, such as a Bonferroni correction. Control of the false positive rate simultaneously increases the false negative rate. Therefore, for analyses of safety, which are most concerned about controlling the false negative rate, it is more conservative not to adjust for multiple testing. The yellow highlighted cells on the summary tab identify those p-values that are less than 0.05. In sensitivity analyses I maintained an overall false positive rate of 0.05 for each AE through a Bonferroni adjustment for the p-values for each AE calculated at the six time periods that I originally considered. This amounts to using a p-value threshold for statistical significance of 0.05/6=0.0083. Because the analysis of the adverse events through April 23, 2004 was a pre-specified analysis of special interest due to a concurrent hold on the Recovery device, and the analyses through February 9, 2006 and June 30, 2006 were also pre-specified due to concurrent events, I did not include them in the joint analysis of the other six time periods and did not adjust the resulting p-values for multiple testing in sensitivity analysis. I additionally applied a Bonferroni correction for the seven time periods that I considered in my January, 2017 analysis; this amounts to using a p-value threshold of 0.05/7=0.007.

Results

The reporting risk ratios and approximate and exact p-values are listed on the summary tab of the Excel workbook.

Analyses of major adverse events:

Filter Embolization Deaths:

By July 2010 [redacted] A reporting risk ratio cannot be calculated due to the denominator of [redacted] but the exact p-value can be calculated and is 8.3e-11, providing evidence of a highly significant increased risk of reports of death due to filter embolization with Recovery filters as compared to SNF filters. This highly significant result is seen starting from Q3'04 for Recovery vs SNF.

Migration:

The reporting risk ratios for migration for [redacted] as [redacted] It is striking that the lower 95% confidence bound for the reporting risk ratio for migration for [redacted] That means that we can be 95% confident that the true reporting risk ratio is at least [redacted] The other lower bounds for the migration risk ratios are [redacted] In all analyses, the reporting risk ratios are all significantly greater than 1, except for [redacted] Similar results are seen when filter embolization is combined with migration ("migration+"), with the exception of [redacted]

Perforation:

[redacted]

Filter Fracture + detached component(s) :

The reporting risk ratios for filter fracture+ are between [redacted]

[redacted]

With respect to the new analysis, I compared the data for the retrievable filters in each time-period with the SNF sales and fracture + detached component reports for November, 2009, because SNF data were not available beyond 2011 and because this time-period contained the highest rate of [redacted] and would amount to a conservative approach as it would use the [redacted] The reporting risk ratios for filter fractures of retrievable devices compared to SNF are [redacted] except for Eclipse in 2010 and Denali in 2013. All of the RRR's [redacted] are statistically significantly different from 1, while those that are [redacted] are not significantly different from 1. There was considerably less data available for Eclipse in 2010 and Denali in 2013, since those were their first years on the market.

Tilted Filter:

The reporting risk ratios are all infinite or undefined due to there being 0 events for [redacted] Statistically significant comparisons of Recovery vs SNF and of G2 vs SNF were found using data through November [redacted] respectively) and through November 2009 [redacted] Comparisons between [redacted] were also significant through November 2009. No new data were reported out to July 2010.

Analyses through specific dates:

In addition to the cumulative analyses described previously, I also considered three specific dates, which are associated with other noted events and issues in which Bard was involved.

Recovery vs SNF through April 23, 2004

The summary table (page 2) of adverse event data through April 23, 2004 (BPVE-01-00268632) [redacted]

G2 vs SNF through February 9, 2006

I have additionally analyzed G2 AE's versus SNF AE's through February 9, 2006. Migration events for G2 through this date are listed in BPVEFILTER-01-00008355. I used the same source for sales data as for the other analyses that I have conducted (BPV-17-01-00193291.xlsx from 2003 and from BPV-17-01-

00188520.XLS for prior to 2003). I was able to infer the migration for SNF through February 9, 2006 using the Bard sheets through Q3 2005 and November 2007 [redacted]

G2 vs SNF through June 30, 2006

I have additionally analyzed G2 AE's versus SNF AE's through June 30, 2006. Migration, perforation and filter fracture events for G2 through this date are listed in 2006.06.30.BPVE-01-01035539_.QUADS R002 native G2 caudal migration failure analysis.pdf (slide 7). I used the same source for sales data as for the other analyses that I have conducted (BPV-17-01-00193291.xlsx from 2003 and from BPV-17-01-00188520.XLS for prior to 2003). I was able to infer the migration and perforation counts for SNF through June 30, 2006 using the Bard sheets through Q3 2005 and November 2007 because they do not change during this bracketing period. [redacted] be conservative, I used the November 2007 count for SNF filter fractures. The reporting risk ratio for G2 vs SNF [redacted]

Analyses of individual filters:

Recovery

[redacted]

G2

G2x and G2/G2x combined

Eclipse

Meridian

Denali

[redacted]

Limitations and responses

There are potential limitations to this analysis that need to be considered when interpreting the results:

- Underreporting: adverse events are generally considered to be underreported to the databases, and potentially differentially by severity of adverse event and by drug or medical device. If only a_1% of occurrences of a particular AE are reported for device 1 and a_2% of occurrences of that AE are reported for device 2, then the RRR will be biased for the true risk ratio, RR, as it will equal $(a_1/a_2)\times RR$. If $a_1=a_2$ then the RRR is unbiased for the RR. If $a_1<a_2$ (i.e., the extent of underreporting for device 1 is greater for device 1 than for device 2), and the true RR is greater than 1, then the RRR is a conservative estimate of the RR (i.e., it is closer to 1 than the RR). It is important to recognize that underreporting in and of itself is not problematic. Rather, differential underreporting of the higher risk device is what leads to bias. And even if there was differential underreporting of the higher risk device, given the variation in reporting relative risks across adverse events, the differential reporting would have had to have been highly variable across adverse events. This does not seem plausible given the severity of the adverse events considered. Given the magnitude of the RRR's, and their variability across adverse events, it seems implausible that differential underreporting by filter could fully explain the deviation of the observed RRR's from 1.

  The estimates of RR provide insight into the magnitude of the underreporting of SNF relative to Recovery that would be necessary to obtain the estimates that were obtained, if the true risk ratio were 1 (i.e., no difference in risk between products). For example, the RRR for Recovery versus SNF for migration events, through July 2010, is 288. This indicates that if there is truly no difference in risk of migration between Recovery and SNF, this reporting risk ratio of 288 could only arise if the probability of reporting a migration event for SNF [redacted]

[redacted]

A concrete instance of underreporting is the lack of [redacted] I handled this by using the number of events and number of sales from the cumulative analysis (through November, 20009) with the largest fracture event rate for SNF. I additionally addressed it in sensitivity analysis by further increasing the [redacted] by adding 5 fractures to the total. However, the true number of events remains unknown [redacted]

In summary, while the reporting risk ratios may involve some degree of underreporting, which makes them imperfect estimates of the actual risk ratios, there is strong evidence across time periods and across AE's of similar degrees of severity to suggest that the true risk ratios are considerably larger than the null value of one.

- Overestimation of denominators: I have used sales data as a proxy for the number of patients who were implanted with the devices. In fact, not every device that was sold was actually implanted in a patient. I addressed this concern by considering in a sensitivity analysis the effect of discounting the sales numbers (through July 2010) by 20%. The reporting risk ratios for this analysis do not change from the original analysis and the p-values are all highly statistically significant ($p<1.0e-7$). If the proportion of filters implanted among those sold does not differ by filter, then this overestimation of exposure does not affect the risk ratio estimation (and would have only negligible effect on the associated p-values). And given that it does not differ by AE prevalence (without differing by filter), this could not explain the observed RRR's given their variance across AE's.

- Counting and data errors: As discussed above, there are discrepancies among company data sheets and even small errors can have large effects on small numbers of events. I addressed this concern by considering in a sensitivity analysis the effect of adding 5 to each of the adverse event totals for SNF in the data through July 2010. The reporting risk ratios are [redacted] Based on my analysis of the Bard AE reports, [redacted]

- No person-time exposure/cannot calculate incidence rates and ratios: Although sales data by month are available, person-time cannot be reliably calculated because it is unknown when devices are removed or replaced. This would be necessary to calculate the incidence rate ratio,

which is useful for individuals as it is a measure of instantaneous relative risk. For analyses at the population level, the risk ratio (which does not take into account exposure time) is a useful measure under some conditions. The use of sales as the denominator does not adjust for time at risk for the AE of interest. This means that estimates of risk are not comparable among products that have different overall person time at risk, unless the risk of the AE is highest close to its implantation and decreasing after that. However, because the SNF was a permanent filter, while Recovery was retrievable, it is reasonable to assume that adjusting for calendar years of sales, there would be greater person time at risk associated with SNF than with Recovery. This would mean that the RRR's present underestimates of the reporting incidence rate ratios, as they multiply them by the ratio of person time exposure of Recovery divided by SNF. Additionally, the adverse events reported for SNF are likely from some implantations that occurred prior to 2000, which are not reflected in the sales numbers that form the denominators of the reporting risk ratios. This means that the estimates for SNF are biased upward and thus the reporting risk ratios are biased downward.

- Temporal effects in reporting ("Weber effect"): Increased reporting can be observed soon after the launch of a drug, and then decrease over time. This does not appear to be the case in this analysis, in [redacted] (BPVE-01-00268632, page 22).

- Reports generated by publicity ("notoriety effect" or "stimulated reporting"): This is known to occur, for example, after FDA warnings appear. It is important to look at timing of the "notoriety effect" to understand what, if any impact, the notoriety has had on reporting. Additionally, if these warnings do not differentially affect the devices, then this does not invalidate the use of the RRR as an estimate of the RR. An FDA warning letter (BPV-17-01-00193337) was sent to Bard on July 13, 2015. It was concerned with manufacturing violations of the Recovery Cone Removal System. It was additionally concerned with violations of Medical Device Reporting requirements, including not reporting to the FDA within 30 days of an event and not submitting complete patient information. While this letter might have generated stimulated reporting, since it was issued in 2015, it would not have any effect on the data used in my analysis. There were no earlier letters regarding concerns about adverse events associated with the Recovery filter, or other Bard products.

- Confounding or "channeling bias": Confounders are patient specific factors, such as age or gender or hypertension, for example, that are associated both with use of a particular device, such as Recovery, and with the AE, such as migration. If the analysis is conducted without adjustment for confounders, the RRR may be biased for the RR. However, if there are patient specific factors associated with use of a particular device, but not with the AE, or conversely, if there are patient specific factors that are associated with the AE, but not with use of a particular

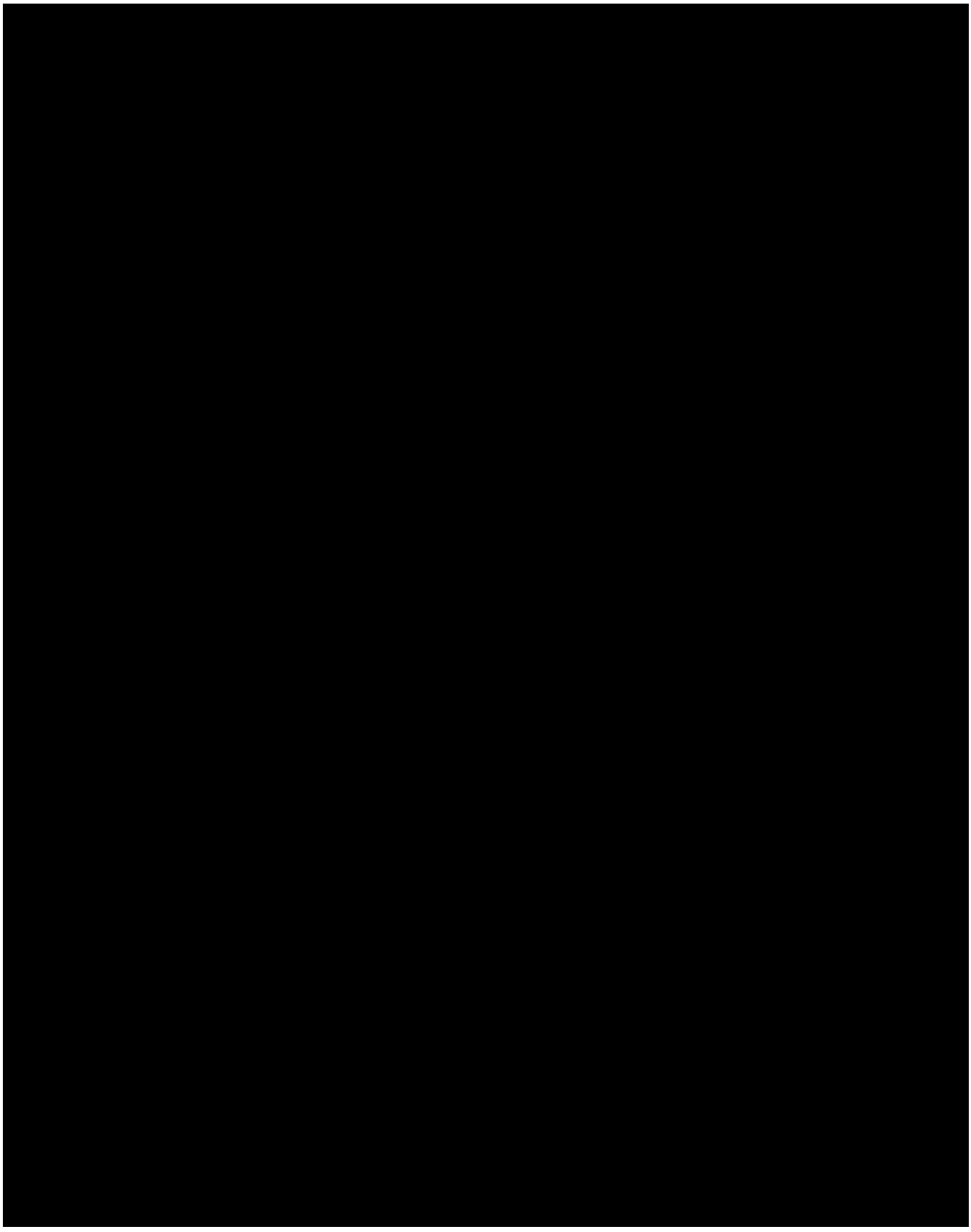
device, the lack of adjustment for them in the analysis does not produce bias. I am not aware of any individual level data that include potential confounding factors.

Summary

[redacted]

In conclusion, the best available adverse event data for the filters considered provide compelling evidence in favor of increased risks of the adverse events that I considered. The reporting risk ratios are generally extremely large, , which in association with their very small p-values, multiple sensitivity analyses, and consistency over time, products and AE's, suggest highly robust [redacted] While these may partially reflect some differential reporting, it is implausible that they would be entirely explained by this.

Filter Migration Test Results:



[redacted]

Digitally signed by Rebecca Betensky
DN: cn=Rebecca Betensky, o, ou, email=betensky@hsph.harvard.edu, c=US
Date: 2017.01.27 15:02:54 -05'00'

Rebecca Betensky, Ph.D.

January 27, 2017

Date